Hearing Research 339 (2016) 195–210

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Capturing contextual effects in spectro-temporal receptive fields

Johan Westö [a,*], Patrick J.C. May [b]

[a] Department of Neuroscience and Biomedical Engineering, Aalto University, FI-00076 Espoo, Finland
[b] Special Laboratory Non-Invasive Brain Imaging, Leibniz Institute for Neurobiology, D-39118 Magdeburg, Germany

ARTICLE INFO

Article history:
Received 30 April 2016
Received in revised form
16 June 2016
Accepted 24 July 2016
Available online 27 July 2016

Keywords:
Spectro-temporal receptive field
STRF
Context field
Inhibition
Synaptic depression
Contextual effects
Auditory

ABSTRACT

Spectro-temporal receptive fields (STRFs) are thought to provide descriptive images of the computations performed by neurons along the auditory pathway. However, their validity can be questioned because they rely on a set of assumptions that are probably not fulfilled by real neurons exhibiting contextual effects, that is, nonlinear interactions in the time or frequency dimension that cannot be described with a linear filter. We used a novel approach to investigate how a variety of contextual effects, due to facilitating nonlinear interactions and synaptic depression, affect different STRF models, and if these effects can be captured with a context field (CF). Contextual effects were incorporated in simulated networks of spiking neurons, allowing one to define the true STRFs of the neurons. This, in turn, made it possible to evaluate the performance of each STRF model by comparing the estimations with the true STRFs. We found that currently used STRF models are particularly poor at estimating inhibitory regions. Specifically, contextual effects make estimated STRFs dependent on stimulus density in a contrasting fashion: inhibitory regions are underestimated at lower densities while artificial inhibitory regions emerge at higher densities. The CF was found to provide a solution to this dilemma, but only when it is used together with a generalized linear model. Our results therefore highlight the limitations of the traditional STRF approach and provide useful recipes for how different STRF models and stimuli can be used to arrive at reliable quantifications of neural computations in the presence of contextual effects. The results therefore push the purpose of STRF analysis from simply finding an optimal stimulus toward describing context-dependent computations of neurons along the auditory pathway.

1. Introduction

An important task for auditory neuroscience is to describe the computations performed by neurons along the auditory pathway. One commonly used model for this purpose is the spectro-temporal receptive field (STRF) (Aertsen et al., 1980; Theunissen et al., 2000; Sharpee, 2013). Traditionally, the STRF assumes that the computations performed by a neuron can be described with a purely linear model (Theunissen et al., 2001; Machens et al., 2004; David et al., 2007) or a linear-nonlinear (LN) cascade model (Chichilnisky, 2001; Lesica et al., 2008; Calabrese et al., 2011), where the linear part in both cases is a similarity measure (dot product) between the spectrogram of the stimulus and the STRF, and where the static nonlinear part describes threshold effects. Both versions therefore represent crude simplifications of neural dynamics, but the resulting STRF provides an easily interpretable description of the computations performed by a neuron.

There are, however, at least two reasons for suspecting that this traditional STRF description is insufficient for describing the computations performed by real neurons: 1) STRFs tend to be dependent on the stimulus so that the STRFs obtained with artificial stimuli, such as ripple combinations, generalize poorly to natural stimuli, such as speech sounds (Blake and Merzenich, 2002; David et al., 2009; Calabrese et al., 2011; for a review see Eggermont, 2011). 2) Contextual effects, giving rise to nonlinear interactions between frequencies or in time, cannot be modeled with a dot product between the spectrogram and the STRF. Examples of such contextual effects are: short-term synaptic plasticity (STP) (Tsodyks and Markram, 1997; Reyes, 2011), basilar-membrane suppression and distortions (Ruggero, 1992; Robles and Ruggero, 2001), and nonlinear facilitation in the frequency and/or time dimension (Sadagopan and Wang, 2009; Brimijoin and O'Neill, 2010; Schneider and Woolley, 2011), where the latter phenomena might emerge out of STP (May and Tiitinen, 2013; May et al., 2015; Westö et al., 2016). In these cases, the nonlinear characteristics of the system mean that the traditional STRF is fundamentally ill-suited

* Corresponding author.
E-mail addresses: johan.westo@aalto.fi (J. Westö), pmay@lin-magdeburg.de (P.J.C. May).

http://dx.doi.org/10.1016/j.heares.2016.07.012
0378-5955/

for describing neural responses. The inability to model nonlinear interactions between spectrogram elements causes a model misfit which may contribute both to the stimulus dependency of STRFs and to STRFs that provide incorrect descriptions of the computations performed by neurons (Froemke and Schreiner, 2015). For example, the observed stimulus dependence of STRFs can be accounted for by STP (David et al., 2009), and STP or nonlinear interactions in general can also give rise to artificial inhibitory regions in STRFs (Ahrens et al., 2008; David and Shamma, 2013), that is inhibitory regions that emerge as a consequence of model misfit.

In order to get accurate descriptions of neural behavior, we therefore need more complex models that can explain the contextual effects exhibited by real neurons, visualize the behavior of these neurons in an easily interpretable way, and predict their responses to arbitrary stimuli. One proposed solution is to estimate several linear filters for the LN model and to make the nonlinearity multidimensional. The estimated filters then span a subspace in stimulus space that is relevant for describing the neuron's behavior (Sharpee et al., 2004; Schwartz et al., 2006; Fitzgerald et al., 2011). However, these filters might be difficult to interpret in terms of a neuron's expected behavior and it is also unknown how contextual effects are manifested in them. A second solution that avoids these problems is to combine the STRF with a context field (CF; Ahrens et al., 2008). The CF models important nonlinear second-order interactions by modifying the stimulus seen by the STRF according to a learned context. The context is neuron-specific and estimated from stimulus/response data together with the STRF. Fig. 1 visualizes the effect of such a learned context for a neuron that responds most strongly to a continuous combination of frequencies, and where the CF hence modifies the spectrogram by highlighting regions where continuous frequencies are active. In practice, CF-equipped models can be thought to operate in two steps: Step 1 extracts a modified spectrogram where each element has been weighted according to context (elementwise multiplication of the original spectrogram with the cross-correlation between the spectrogram and the CF). Step 2 determines the response using a (possibly nonlinear) firing function and the similarity score (dot product) between the modified spectrogram and the STRF. The second step is equivalent to the traditional definition of the STRF in the LN-model framework, the only difference being that the STRF is now compared to a context-processed version of the spectrogram. The CF hence extends STRF analysis to also include cases where nonlinear interactions are present. It is applicable in situations where traditional STRF analyses have been used previously, and it might even provide new information in these situations as possible contextual effects can be verified and visualized in the CF.

In real life, true STRFs are always unknown. It is therefore very difficult to know whether any particular STRF analysis suffers from model misfit and whether including a CF actually provides a truer description of the STRF. These questions can only be answered through simulations, where the input, structure, and dynamics of the neural system are fully known. Unfortunately, STRF models are normally only evaluated on simulated data from the LN model equipped with one or more linear filters (Paninski, 2004; Sharpee et al., 2004; Fitzgerald et al., 2011; Meyer et al., 2014a; but see Christianson et al., 2008), and hence we do not know how these models perform when the real neural dynamics exhibits any type of contextual effects. In the current study, we addressed this problem by: 1) defining the extended STRF, a nonlinear extension of the traditional STRF that also lets us describe true STRFs for neurons exhibiting contextual effects, and 2) setting up a simulated environment for testing a wide range of STRF models to differences in stimuli and the presence/absence of contextual effects.

2. Materials and methods

2.1. A nonlinear extension of the traditional STRF

Traditionally, the STRF has been defined as a purely linear model (Theunissen et al., 2001; Machens et al., 2004; David et al., 2007) or as the linear part in an LN model with one filter (Chichilnisky, 2001; Lesica et al., 2008; Calabrese et al., 2011). In the slightly more complex LN interpretation, the computations performed by a neuron are described with a similarity score (z) and a nonlinearity $f(\cdot)$ as:

$$\begin{aligned} z_i &= w_0^{\mathrm{strf}} + \sum_j^{N_{\mathrm{strf}}} w_j^{\mathrm{strf}} x_{ij}^{\mathrm{strf}}, \\ y_i &= f(z_i), \end{aligned} \tag{1}$$

where w_0^{strf} is a bias term, $\mathbf{w}^{\mathrm{strf}}$ the parameter vector representing the concatenated STRF, $\mathbf{x}_i^{\mathrm{strf}}$ the concatenated input vector (stimulus spectrogram), and N_{strf} is the number of elements in the STRF. However, this description does not allow nonlinear interactions to be modeled, and hence, it is not possible to define a true STRF for a neuron in cases where nonlinear interactions between spectrogram elements are needed to describe the contextual effects present. We therefore define a nonlinear extension to the traditional STRF by modifying the similarity score as:

$$z_i = w_0^{\mathrm{strf}} + \sum_j^{N_{\mathrm{strf}}} w_j^{\mathrm{strf}} x_{ij}^{\mathrm{strf}} g_j\left(\mathbf{x}_{ij}^{\mathrm{ctx}}\right), \tag{2}$$

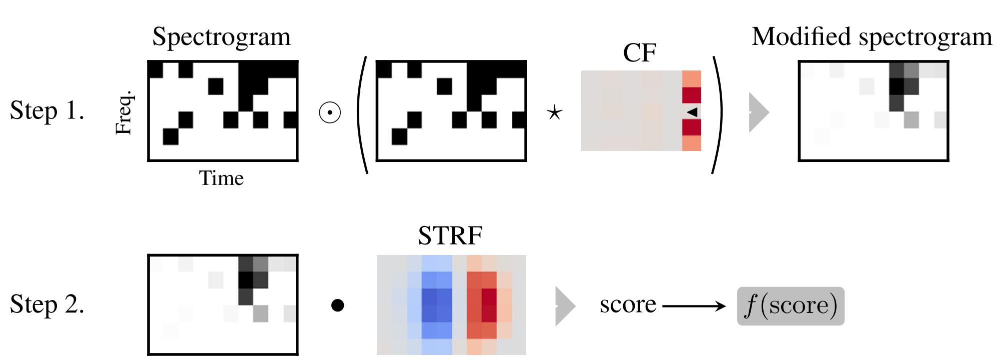

Fig. 1. CF-equipped STRF models can describe contextual effects by modeling important second-order nonlinearities. Such models can be thought to operate in two steps: Step 1 extracts a modified spectrogram through elementwise multiplication ($\odot$) between the original spectrogram and a cross-correlation ($\star$) between the original spectrogram and the CF (◄ indicates the origin of the CF). Step 2 determines a similarity score between the modified spectrogram and the STRF as a dot product ($\bullet$) and sends the result through a (possibly nonlinear) firing function (f) to obtain a predicted response.